

For discussion purposes only

DO NOT ENTER /KRS/ 06 29/2009

Ref. 10/797,359

Hello Examiner Skowronek,

Regard our conversation Tuesday, thank you for offering a review of proposed amendments to help bring this case to completion. Please find the following amendments which I hope you can comment on regarding possible allowability in your opinion. Note these amendments are not meant acknowledge any agreement by the Applicant with the rejections.

Referencing the amendments it appears that none of the references can be seen to partition text to recognize chemical name fragments using dictionaries consisting of a common chemical prefix dictionary and a common chemical suffix dictionary. In paragraph [0039] of the published application it is explicitly stated that the invention uses "two small dictionaries."

Also I can not find where in any of the references it can be found to disclose or suggest operations related to "chemical name fragments and the substructures that do not contain a number." The related amendments are supported at least in paragraphs [0074] – [0075] of the published application.

Then also, I can not find in the references where Hull relates to "where the search comprises first entering one or more chemical fragment names and then entering one or more substructures in a representation form, where the entering is by at least one of text form or graphical selection."

Also please note amendments to **claims 8, 10, 13, 15, and 16-17** which may be incorporated if approved.

Your responses regarding these amendments or other amendments you feel may provide an allowance are greatly appreciated.

Thank you very much,
John Garrity
203-925-9400 (x39).

1. (Currently Amended) A method to process a text document, comprising:

partitioning text of the text document and assigning semantic meaning to words of the partitioned text, where assigning comprises applying a plurality of regular expressions, rules and a ~~plurality of~~ dictionaries consisting of a common chemical prefix dictionary

For discussion purposes only

and a common chemical suffix dictionary to recognize chemical name fragments;

recognizing any substructures present in the chemical name fragments;

extracting information associated with the recognized chemical name fragments and the substructures of the text document and indexing the extracted information in a text index;

adding each of the recognized chemical name fragments and the substructures that do not contain a number to the text index;

determining structural connectivity information of each of the recognized chemical name fragments and recognized the substructures that do not contain a number;

~~extracting information associated with the recognized chemical name fragments and substructures of the text document and indexing the extracted information in a text index;~~

indexing representations of the recognized chemical name fragments and the substructures in association with the determined structural connectivity information into a plurality of chemical connectivity tables of a chemical substructure index;

storing the text index in association with the ~~indexed representations~~ chemical substructure index ~~in a searchable index;~~ and

For discussion purposes only

providing a graphical user interface to search the ~~searchable~~ text index and the chemical substructure index, where the search comprises first entering one or more chemical fragment names and then entering one or more substructures in a representation form, where the entering is by at least one of text form or graphical selection.

2. (Currently Amended) A The method as in claim 1, wherein the extracting further comprises extracting keywords from the text document and indexing the keywords in the text index, and wherein the search comprises selecting a graphical representation of one or more substructures and additionally entering at least one keyword.

3. (Currently Amended) A The method as in claim 1, wherein extracting further comprises extracting keywords from the text document and indexing the keywords in the text index, and wherein the search comprises additionally entering at least one keyword, and at least one of chemical name fragment connectivity and substructure connectivity.

4. (Canceled)

5. (Canceled)

6. (Currently Amended) A The method as in claim 1, wherein the search further comprises entering at least one search term, and where a search results in an intersection of the ~~indexed representations~~ chemical substructure index and the text index, identifying at least one document where there are found chemical compounds that contains a

For discussion purposes only

reference to a corresponding chemical compound the one or more substructures and the at least one search term.

7. (Currently Amended) A The method as in claim 1, where determining structural connectivity information comprises looking up recognized chemical name fragments and substructures in a structure dictionary.

8. (Currently Amended) A The method as in claim 1, where the indexing representations comprise MOL type representations and SMILES type representations comprises:

testing if each of the recognized chemical name fragments occur in a SMILES fragment dictionary, where if it does occur in the SMILES fragment dictionary then adding the chemical name fragment to the chemical substructure index as the SMILES representation, and

testing if each of the recognized chemical name fragments occur in a MOL file fragment dictionary, where if it does occur in the MOL file dictionary then adding the chemical name fragments to the chemical substructure index as the MOL file representation.

9. (Currently Amended) A The method as in claim 1, where said plurality of dictionaries comprise consists of a dictionary of common chemical prefixes and a dictionary of common chemical suffixes.

10. (Currently Amended) A The method as in claim 1, where said plurality of dictionaries comprise consists of a dictionary of common chemical prefixes, a dictionary of common

For discussion purposes only

chemical suffixes, and a dictionary of stop words to eliminate erroneous chemical name fragments.

11. (Currently Amended) A The method as in claim 1, further comprising filtering recognized chemical name fragments using a list of stop words to eliminate erroneous chemical name fragments.

12. (Currently Amended) A The method as in claim 1, where chemical name fragments are further recognized by using common chemical word endings.

13. (Currently Amended) A The method as in claim 1, where application of said regular expressions and rules results in punctuation characters being ~~one of maintained or removed from~~ between chemical name fragments as a function of context.

14. (Currently Amended) A The method as in claim 1, where said regular expressions comprise a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation.

15. (Currently Amended) A The method as in claim 14, where the punctuation comprises ~~at least one of~~ a parenthesis, a square bracket, a hyphen, a colon and a semi-colon.

16. (Currently Amended) A The method as in claim 14, where the characters comprise ~~at least one of~~ upper case C, O, R, N and H.

For discussion purposes only

17. (Currently Amended) A The method as in claim 14, where the characters comprise strings of ~~at least one of~~ lower case xy, ene, ine, yl, anc and oic.

18. (Currently Amended) A The method as in claim 1, comprising an initial step of tokenizing the document to provide a sequence of tokens.